Nucleic Acids Research, 1999, Vol. 27, No. 2 455–461

Pogo transposase contains a putative helix–turn–helix DNA binding domain that recognises a 12 bp sequence within the terminal inverted repeats

Hongmei Wang[+], Eve Hartswood and David J. Finnegan*

Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK

Received October 13, 1998; Revised and Accepted November 19, 1998

ABSTRACT

***Pogo* is a transposable element with short terminal inverted repeats. It contains two open reading frames that are joined by splicing and code for the putative *pogo* transposase, the sequence of which indicates that it is related to the transposases of members of the *Tc1/mariner* family as well as proteins that have no known transposase activity including the centromere binding protein CENP-B. We have shown that the N-terminal region of *pogo* transposase binds in a sequence-specific manner to the ends of *pogo* and have identified residues essential for this. The results are consistent with a prediction that DNA binding is due to a helix–turn–helix motif within this region. The transposase recognises a 12 bp sequence, two copies of which are present at each end of *pogo* DNA. The outer two copies occur as inverted repeats 14 nucleotides from each end of the element, and contain a single base mismatch and indicate the inverted repeats of *pogo* are 26 nucleotides long. The inner copies occur as direct repeats, also with a single mismatch.**

INTRODUCTION

Transposable elements that have short terminal inverted repeats, known as transposons or Class II elements (1), generally move by an excision–insertion mechanism and encode proteins, transposases, required for their own transposition. During transposition the ends of an element must be brought together, released from flanking sequences and joined to target DNA. These steps are mediated by transposases acting alone or in concert with other proteins. In carrying out these steps, a transposase must bring together the ends of an element before cleaving both strands of DNA at the junction between the element and sequences flanking it. This is the result of a nucleophilic attack on the appropriate phosphodiester bond with water acting as the nucleophile (2). After strand cleavage the 3′OH groups produced at each end of the transposon attack target DNA in *trans*-esterification reactions that covalently link each strand of the transposon to the target (3). If the sites at which the target is cleaved are staggered, as appears always to be the case, then there will be a short single-stranded region at the junction between the transposon and the flanking DNA. Repair of these regions by host factors generates target site duplications. These reactions have been studied *in vitro* using the transposases of the *Tc1* and *Tc3* elements of *Caenorhabditis elegans* (4,5), the *P* element of *Drosophila melanogaster* (6) and *Himar1* of *Haemotobia irritans* (7).

Pogo is a transposable element of this type. It was first discovered as a 190 bp insertion associated with the *white-eosin* mutation in *D.melanogaster* (8). This was used as a probe with which to investigate the size of *pogo* elements in DNA from different strains of *D.melanogaster*. Most strains have many copies of the 190 bp element, 10–15 copies of 1.1–1.5 kb elements and several copies of a 2.1 kb element that is believed to be intact. One of these full-length elements has been sequenced and is 2121 bp long with 21 bp terminal inverted repeats and two open reading frames (ORFs). The smaller elements appear to be derived from the largest by single internal deletions as all elements have the same termini (8).

Analysis of *pogo* cDNAs indicates that the two ORFs of the longest elements are joined by RNA splicing to encode a single polypeptide of 499 residues if initiated at the first methionine codon (8). This is presumed to be the *pogo* transposase and is similar to the putative transposases of *Fot1* and related elements in fungi (9,10), *Tigger* elements in the human genome (11,12) and *Tc2* (13), *Tc4* and *Tc5* of *C.elegans* (14,15). *Pogo* transposase is also similar to some proteins that are not transposases including the mammalian centromere binding protein CENP-B (16), the jerky protein involved in epileptic seizures in mice (17), and the yeast regulatory proteins RAG3 and PDC2 (18,19).

The majority of eukaryotic transposons identified so far encode proteins related to the transposases of *mariner* elements of *Drosophila* and *Tc1* elements of *C.elegans* (20,21). These can be recognised because they share a presumed catalytic domain that includes two aspartate residues separated by ~90 amino acids followed by a third acidic residue that is 34 or 35 amino acids downstream. This motif is known as D,D35E. This is also found in the transposases of some bacterial transposable elements and in retroviral integrases (3). *Pogo* appears to be a distant member of this super-family of elements but the putative transposase that it encodes has no clear candidate for the last acidic residue of the D,D35E motif (12).

*To whom correspondence should be addressed. Tel: +44 131 650 5361; Fax: +44 131 650 8650; Email: d_finnegan@ed.ac.uk

[+]Present address: Department of Microbiology, Eastman Research Institute, 256 Gray's Inn Road, London WC1X 8LD, UK

The ability of a transposase to recognise specifically the ends of its own element is essential for transposition. The transposases of *Tc1* and *Tc3* have bipartite DNA binding domains near their N-termini. The first part is responsible for recognition of sequences within the terminal inverted repeats of the element concerned. No sequence-specific binding activity has been found for the second DNA binding domain but it may interact with DNA close to the cleavage site (22–24). These transposases are predicted to have helix–turn–helix (HTH) motifs within each of these DNA binding domains, whereas the transposases of *mariner* and *pogo* are predicted to have a single HTH (25). The crystal structure of the first 65 residues of *Tc3* transposase together with the DNA that it recognises has confirmed that this region of the protein contains an HTH and that this is responsible for DNA binding (24). The transposase of bacteriophage *Mu* (MuA protein) contains two HTH regions at its N-terminus, each of which can recognise the 22 bp transposase binding sites, three copies of which are present at each end of *Mu* DNA (26,27).

In order to investigate the DNA binding properties of the putative *pogo* transposase and to determine the sequence that it recognises we have expressed the protein in *Escherichia coli* as a fusion with glutathione-S-transferase (GST). We have shown that this protein binds specifically to sequences at each end of *pogo* and that the region of the protein responsible for DNA binding is within its N-terminal 74 amino acids, the part of the protein that has been predicted to contain an HTH motif (25). Mutations that change specific residues within the second helix of the HTH greatly reduce DNA binding by this fusion protein, indicating that the HTH is required for sequence-specific recognition of the terminal inverted repeats by *pogo* transposase. We have also identified a 12 bp sequence that is recognised by the HTH.

MATERIALS AND METHODS

The sequences of the oligonucleotide primers mentioned below can be obtained from http://www.icmb.ed.ac.uk/research.html#Finnegan

Construction of expression plasmids

A plasmid for expression of full-length transposase was constructed as follows. Primers M0804 and M0805 are complementary and contain the sequence at the junction of ORF1 and ORF2 of *pogo* as indicated by the sequence of *pogo* cDNAs (8). Primers S5046 and S5049 comprise the sequences at the beginning of ORF1 and end of ORF2, respectively. In S5046 the ORF1 sequence is preceded by a *Bam*HI site and in S5049 the ORF2 sequence is followed by an *Xho*I site. The sequences of ORF1 and ORF2 were amplified separately using primers S5046 and M0804 and primers M0805 and S5049 and a template from pogoR11XC (8). The products of these reactions were mixed in equimolar amounts and used as templates for a second PCR using primer S5046 and S5049 as primers. The final PCR product was purified by GENECLEAN (BIO101) and cloned into a pGEX4T-2 vector (Pharmacia) via *Bam*HI and *Xho*I sites in a ligation reaction catalysed by T4 ligase (New England Biolabs). This joins the C-terminus of GST to the second codon of transposase omitting the methionine that is presumed to initiate translation. Competent *E.coli* NM522 cells were then transformed and the transformants were selected on L-Amp plates (100 μg/ml ampicillin). The positive constructs were examined by DNA sequencing before being used for protein expression.

DNA coding for various fragments of *pogo* transposase were amplified by PCR using different pairs of primers: S5046 and A05 for N158, S5046 and A06 for N138, S5046 and T6395 for N74, S5046 and T5774 for N59, Z7187 and S5049 for C430, A09 and S5049 for C380, A10 and S5049 for C341, and S5048 and S5049 for C193; these PCR products were cloned into pGEX4T-2 using the same restriction sites as the whole transposase as described above.

Pairs of complementary primers were designed to generate single amino acid substitutions within GST:N74 as follows: V7017 and V7018 for the R39A mutation, V7013 and V7014 for the R44A mutation, V7015 and V7016 for the mutation K48A, W0914 and W0915 for K34A, W0916 and W0917 for R63A, V7912 and V7913 for C34P, and V7914 and V7915 for V42P. The minus strand primers were used together with S5046 to amplify the left hand part of the mutated N74 proteins, while the plus strand primers were used together with T6395 to amplify the right hand part. S5046 and T6395 were then used to join the left and right hand segments of the DNA coding for the proteins using equimolar amounts of the corresponding left and right hand PCR products as templates. The products of these reactions were cloned into pGEX4T-2 as described above for the complete transposase coding sequence.

Protein expression and purification

Each transposase derivative was expressed in *E.coli* NM522 (28). Expression of the fusion protein was induced at OD_{550} by adding 1 M IPTG to a final concentration of 0.5 mM followed by incubation at 30°C for 2–5 h. Cells from 1.5 ml of culture were resuspended in 300 μl of ice-cold PBS (80 g NaCl, 20 g KCl, 20 g Na_2HPO_4, made up to 1 l with H_2O, pH 7.4) and 1% (v/v) Triton X-100. The cells were lysed by sonication and the insoluble material removed by centrifugation for 5 min at 4°C. The supernatant was then transferred to a fresh tube together with 50 μl of a 50% slurry of glutathione-agarose beads. This was incubated at 4°C for 0.5–2 h to allow binding of the GST fusion proteins. The beads were then collected by centrifugation (5 s at 14 000 *g*) and washed three times with 1 ml of PBS. If the fusion protein was to be analysed by SDS–PAGE the beads were incubated at 100°C for 3 min in loading buffer with 6% SDS and 15% (v/v) β-mercaptoethanol. If the protein was to be used for DNA binding studies, it was eluted using an equal volume of freshly made 50 mM Tris–HCl (pH 8.0) containing 5 mM reduced glutathione (Sigma) pH 7.5.

Preparation of probes for gel retardation assays

A DNA fragment to be labelled was amplified by PCR and digested with restriction enzymes to create a cohesive end at either one or both ends. Different pairs of primers were used to amplify different DNA fragments as follows: A02 and A01 for the 131 bp probe, A02 and T6396 for the 95 bp probe, A02 and T7175 for the 43 bp probe, A03 and A04 for the 22–160 probe, T7880 and A01 for the 43–131 probe, V5498 and T6396 for the –25–95 probe, V5498 and T7461 for the –25–43 probe, V5498 and V6333 for the –25–21 probe, and V8599 and V8600 for the R12 probe. Primers V5498, V7460, V7459 and T6396 were used to generate the internally deleted probe –25–95(Δ22–42) using a method equivalent to that described above for joining the two coding regions of the transposase. Oligonucleotides V8601 and V9045 were annealed to make the 14BS probe.

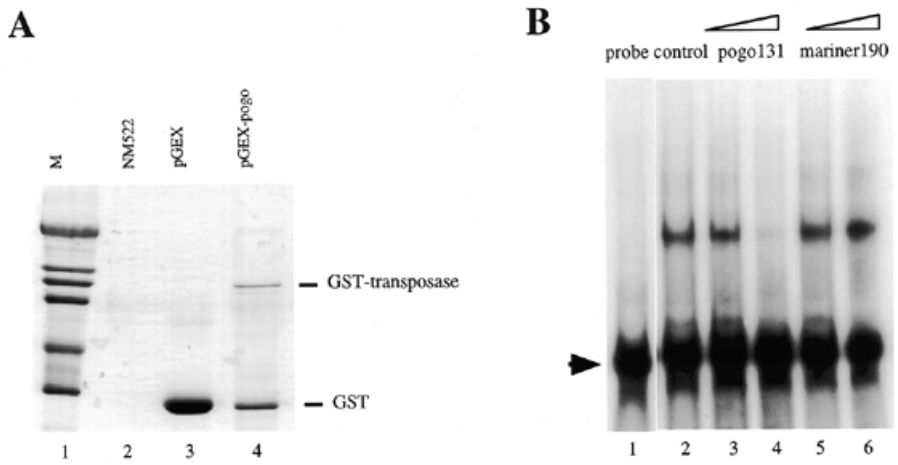

Figure 1. (**A**) Purification of GST-transposase fusion protein. Extracts of *E.coli* strain NM522 with or without an expression plasmid and grown in the presence of IPTG were mixed with glutathione-agarose beads. Proteins released from the beads by boiling in the presence of SDS were separated on a 12% SDS polyacrylamide gel and stained with Coomassie Blue. Lane 1, marker proteins (Sigma SDS-6H) of 29, 45, 66, 97, 116 and 205 kDa; lane 2, protein from cells with no plasmid; lane 3, protein from strain NM522 carrying the vector pGEX4T-2; lane 4, proteins from strain NM522 carrying the fusion plasmid pGEX-pogo. (**B**) *Pogo* transposase binds to the left hand end of *pogo* in a sequence-specific manner. A fragment comprising the first 131 bp of *pogo* was labelled with ^{32}P and incubated with purified GST:transposase with or without unlabelled competitor DNA before being separated on a 5% polyacrylamide gel. The position of the probe fragment is indicated by an arrowhead. Lane 1, probe alone; lanes 2–6, probe plus GST:transposase; lanes 3 and 4, incubation with a 10- or 100-fold molar excess of unlabelled probe fragment; lanes 5 and 6, incubation with a 10- or 100-fold molar excess of a 190 bp fragment from the left hand end of the *mariner* transposable element.

DNA fragments to be used as probes were radioactively labelled with [α-^{32}P]dCTP using Klenow polymerase (New England Biolabs) to fill in cohesive ends generated by digestion with restriction enzymes.

Gel retardation assay

An aliquot of 1–3 µg of protein was incubated with 1 µg of non-specific competitor poly dI-dC on ice in binding buffer (25 mM HEPES, pH 7.6; 40 mM KCl; 2 mM $MgCl_2$; 0.1 mM EDTA; 1 mM DTT; 10% glycerol) for 10 min. DNA probe (2 ng) was added to the mixture and the incubation continued for another 20 min. In competition assays a 10–100-fold molar excess of unlabelled competitor DNA was added to the reaction at the same time as poly dI-dC. Two microlitres of loading buffer (binding buffer with 0.05% bromophenol blue) was added to each sample before it was run on a 5% polyacrylamide gel. The gel was then dried under vacuum and autoradiographed.

RESULTS

Sequence-specific DNA binding activity of *pogo* transposase

In order to investigate the DNA binding properties of *pogo* transposase we have expressed the protein in *E.coli* as a fusion with GST. The two exons of the transposase gene were joined according to the sequence of *pogo* cDNAs (8) using PCR as described in Materials and Methods. The complete transposase coding sequence, except for the initiating methionine, was then inserted downstream of the GST gene in the vector pGEX4T-2 (Pharmacia) to give the plasmid pGEX-pogo. This plasmid was introduced into the *E.coli* strain NM522 and GST-transposase expression induced by the addition of IPTG after which the fusion protein was purified on glutathione-agarose beads (Materials and Methods). A protein of about the expected size, 85 kDa, was produced from cells carrying pGEX-pogo but not from those carrying unmodified pGEX4T-2 (Fig. 1A). The fusion protein contains a thrombin cleavage site at the junction between GST and transposase. A polypeptide corresponding in size to GST itself was isolated from cells carrying pGEX-pogo. This is presumably the result of premature termination of translation or proteolytic cleavage at the junction between the GST and transposase sequences.

Transposases are believed to recognise the ends of their own elements as the first step in transposition to form synaptic complexes before catalysing endonuclease cleavage and strand transfer reactions. Sequence-specific DNA binding should therefore be a characteristic of every transposase. Since the sequences recognised by transposases studied so far lie within the terminal inverted repeats of an element (7,23,29) or sequences adjacent to them (6,30) we have used a probe containing 131 bp from the left hand end of *pogo* as a probe with which to investigate the DNA binding properties of its transposase in gel retardation assays.

The purified GST-transposase fusion protein bound the end probe in a sequence-specific manner as binding was competed by a 100-fold molar excess of unlabelled probe fragment but not by a similar amount of a 190 bp fragment from the left hand end of the transposable element *mariner* when it was used as a non-specific competitor (Fig. 1B). No binding of the *pogo* probe was seen when it was incubated with either purified GST itself or with whole cell extracts of the host strain NM522 (data not shown).

The DNA binding domain of *pogo* transposase

We have determined the region of *pogo* transposase that is responsible for DNA binding by making fusion proteins with GST and various C- or N-terminal deletions of the transposase and testing their ability to bind to the 131 bp probe. The C-terminal deletions resulted in fusion proteins containing the N-terminal 158, 138, 74 and 59 amino acids of the transposase, respectively, and of these all but the last were able to bind the

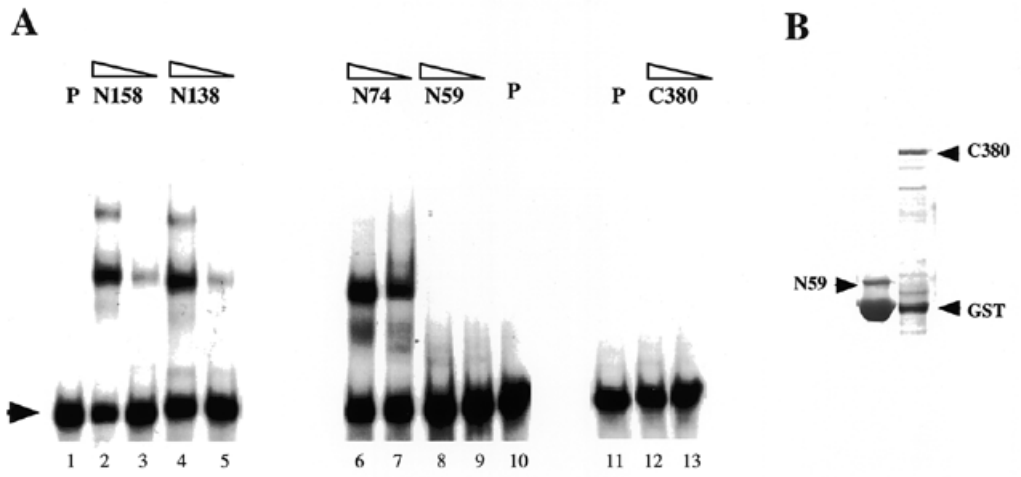

Figure 2. The DNA binding domain of *pogo* transposase is contained within the N-terminal 74 residues. (**A**) A fragment comprising the left hand 131 bp of *pogo* was labelled with ^{32}P and incubated with 1 or 3 µg of GST-transposase fusion proteins and then separated on a polyacrylamide gel. Lanes 1, 10 and 11 show the probe alone; lanes 2–5, 6–9, 12 and 13 show the probe incubated with GST fused to segments of *pogo* transposase. The length of the transposase sequence is indicated above each lane. Lanes 2, 4, 6, 8 and 12 show the effect of incubating with 1 µg of protein and lanes 3, 5, 7, 9 and 13 of incubating with 3 µg. 'N' indicates residues from the N-terminus while 'C' indicates residues from the C-terminus. The position of the probe fragment is indicated by an arrowhead. (**B**) SDS–PAGE gel showing the GST-transposase fusions N59 and C380 used for the gel retardation experiments in (A). The protein concentrations were adjusted appropriately prior to incubation with the probe.

131 bp probe (Fig. 2, lanes 1–10). This suggests that a DNA binding domain of *pogo* transposase is located at the N-terminus of the protein with at least one end lying between residues 59 and 74. None of the fusion proteins containing N-terminal deletions of the transposase sequence bound the probe (Fig. 2, lanes 11–13). Since the longest of these, GST:C430, contained all but the N-terminal 70 residues, we conclude that the DNA binding capacity of *pogo* transposase is due to a sequence, or sequences, within the N-terminal 74 residues of the protein.

In gel retardation experiments using GST fused to the full-length transposase (Fig. 1), and with GST:N158 and GST:N138 (Fig. 2, lanes 2 and 4), two retarded bands can be seen, the larger being much less prominent than the smaller. These higher molecular weight bands are only seen at higher protein concentrations (Fig. 2, lanes 2–5) and may be due to the binding of transposase dimers. If dimers are formed then the dimerisation domain is probably downstream of the DNA binding domain, since a single retarded band was seen with GST:N74 even at high protein concentrations. The fact that only a single retarded band was seen with GST:N74 suggests that the additional bands seen with other transposase fusions were not due to transposase monomers binding separately to the two binding sites on the probe. It is also unlikely that these minor bands are due to degradation of the transposase fusion proteins, since full-length proteins have always been the major species in our preparations. This can be seen for the full-length protein in Figure 1A, lane 4.

Pietrokovsky and Henikoff have predicted that *pogo* transposase contains an HTH motif from residues 26 to 47 (25). This lies within the region that we have shown is responsible for DNA binding (Fig. 3A), so we have used site-directed mutagenesis to determine whether or not the residues predicted to form this HTH are required for binding. The residues in an HTH that make contact with the DNA sequence that it recognises are generally in the second, or recognition, helix (31), so we have investigated the effect on DNA binding of changing arginine 35 and arginine 43 to alanine residues (R39A, R44A) (Fig. 3A) in the fusion protein GST:N74. The mutant proteins were expressed in *E.coli* and purified on glutathione-agarose beads before being tested for DNA binding with the 131 bp *pogo* probe (Fig. 3B, lanes 4–7). The ability of GST:N74 to bind to the *pogo* probe was greatly reduced when either of the residues believed to lie in the second helix of the HTH motif was changed, whereas changing the lysine residue just beyond the second helix (K48A) had little effect (Fig. 3B, lanes 8 and 9). We have also made equivalent changes to lysine 34 (K34A) in the putative first helix of the HTH and to arginine 63 (R63A) beyond the recognition helix (Fig. 3A). Neither of these mutations had a significant effect on DNA binding (data not shown).

We have investigated the importance of this region of the protein further by introducing proline residues separately into each potential helix of the HTH (C32P and V42P, Fig. 3A). The resulting GST:N74 derivatives also had less affinity for the probe than the wild-type and the retarded complexes that were formed migrated more slowly on the gel (Fig. 3B, lanes 10–13), possibly because of the effect of the proline residues on their conformation.

These results are consistent with the prediction that the ability of *pogo* transposase to bind ends of the element is due to an HTH motif lying between residues 28 and 47 (25).

Binding site for *pogo* transposase

In order to identify the sequence recognised by *pogo* transposase we have tested the ability of GST:N74 to bind to probes comprising different sequences from the left hand end of the element (Fig. 4). Deletion of DNA between nucleotides 43 and 131 did not affect binding (Fig. 5, lanes 1–3) whereas a probe containing nucleotides 43–131 was not bound by transposase (Fig. 5, lanes 10–12) indicating that the sequence recognised by transposase lies within the first 43 bp. This is consistent with our expectation that the transposase binding site is contained within

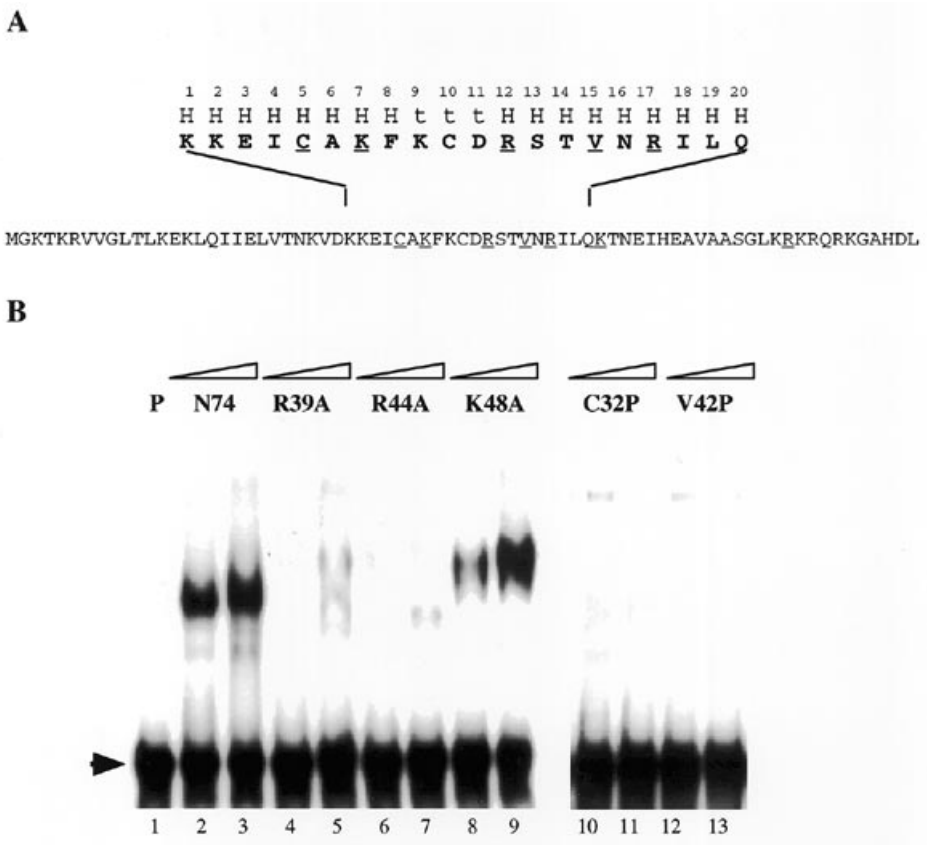

Figure 3. The N-terminus of *pogo* transposase may contain an HTH motif. (**A**) The residues believed to be contained within a helix (H) or turn (t) are indicated above the sequence of the N-terminal 74 residues of *pogo* transposase. The residues that have been altered by site-directed mutagenesis are underlined. The top line indicates the numbering of residues within the predicted HTH while the bottom line indicates the numbering of the residues within the transposase. (**B**) The effect on DNA binding of changes to specific residues in the DNA binding domain of *pogo* transposase. The 131 bp probe from the left hand end of *pogo* was incubated with wild-type or mutant GST:N74 as indicated, before being separated on a polyacrylamide gel. P indicates the probe alone. The position of the probe fragment is indicated by an arrowhead.

the terminal inverted repeat as is the case for *Tc1*, *Tc3* and *Himar1* (7,23). This is not the case, however, as a 46 bp probe containing the 21 bp inverted repeat was not bound by transposase (Fig. 5, lanes 7–9), whereas a probe containing nucleotides 22–161 was bound (Fig. 5, lanes 4–6). The 46 bp probe contained 25 bp of non-*pogo* sequence to the 5′ side of nucleotides 1–21 in case transposase binds less efficiently to short fragments of DNA. We conclude that binding requires nucleotides 22–43 and have confirmed this by showing that a probe containing nucleotides 1–95, but deleted for nucleotides 22–43 (data not shown), is not bound by transposase even though the equivalent 1–95 bp probe is bound (Fig. 4).

Transposase must recognise both ends of *pogo* during transposition and as the 21 bp terminal inverted repeats at each end are identical we expected that transposase would bind to a subterminal sequence at the right hand end. We have tested this using a probe comprising the 140 nucleotides immediately to the left of the right hand inverted repeat (nucleotides 1960–2100). This was bound by transposase in a gel retardation assay, indicating that it contains at least one transposase binding site (data not shown).

We have determined the nucleotide sequence recognised by transposase by comparing nucleotides 22–43, which contain at least one transposase binding site, with the sequences at each end of *pogo*. This should identify any sequence that is also present at the right hand end as well as any sequence that occurs more than once at the left hand end. This revealed a 12 bp sequence that is present in inverted orientation and with one mismatch. These lie

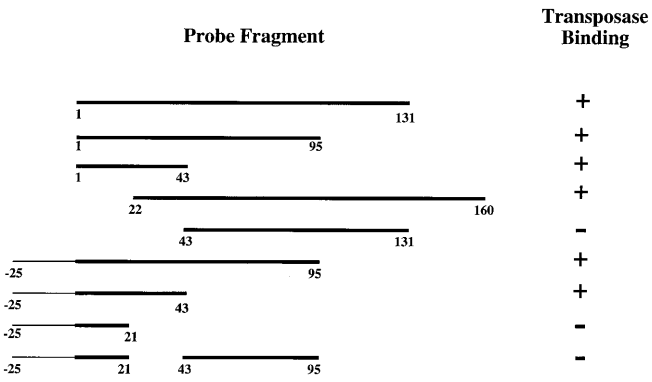

Figure 4. The position of the sequences recognised by *pogo* transposase. The ability of the GST:N74 fusion protein to bind to fragments comprising various sequences from the left hand end of *pogo* is indicated. The 25 nucleotides to the left of the left hand end of the *pogo* sequence (–25 to 1) is the chromosomal sequence immediately to the left of the *pogo* element R11 (8).

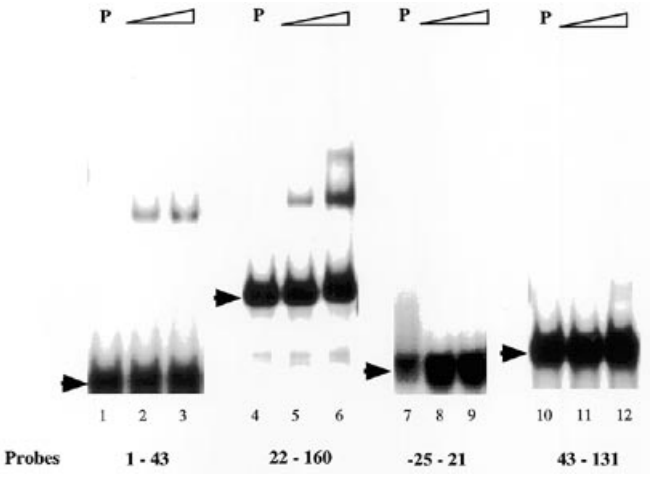

Figure 5. Interaction of the DNA binding domain of *pogo* transposase with sequences from the left hand end of *pogo*. Increasing amounts (0, 1 and 3 µg) of the GST:N74 fusion protein were incubated with ^{32}P-labelled probes comprising sequences from the left hand end of *pogo* as indicated. The positions of the free probe fragments are indicated by arrowheads. The probes correspond to the fragments shown in Figure 4 as follows: lanes 1–3, nucleotides 1–43; lanes 4–6, nucleotides 22–160; lanes 7–9, nucleotides –25 to 21; lanes 10–12, nucleotides 43–131.

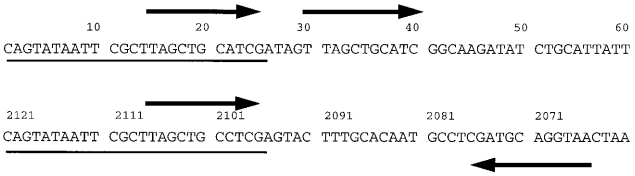

Figure 6. The position of transposase binding sites at the ends of *pogo*. The upper sequence is from the top strand at the left hand end of *pogo* and the lower sequence is from the bottom strand at the right hand end. Both sequences are written 5′ to 3′. The positions of the 12 bp transposase binding sites are indicated by arrows. The 26 bp terminal inverted repeats are underlined.

14 bp from each end of *pogo* [nucleotides 14–25 (TTAGCTG-CATCG) and 2108–2097 (TTAGCTGCCTCG)]. This extends the terminal inverted repeats, previously thought to be 21 bp long, past a single base mismatch to a length of 26 bp (Fig. 6).

We have confirmed that both copies of the 12 bp sequence are recognised by transposase by showing that GST:N74 binds to both a 43 bp synthetic oligonucleotide containing the 12 bp sequence flanked by sequences not present at the ends of *pogo* (Fig. 7, lanes 1–3) and a 43 bp oligonucleotide comprising nucleotides 2079–2121 from the right hand end of *pogo* (Fig. 7, lanes 4–6). Copies of the 12 bp sequence are also present at asymmetric sites just inside the inverted repeats, an exact copy at nucleotides 30–41 and a copy with one mismatch (TTAC-CTGCATCG) at nucleotides 2066–2077 (Fig. 6). The first of these can account for the binding of GST:N74 to nucleotides 22–43 (Fig. 4). DNA containing the 12 bp sequence from nucleotides 2066–2077 is also recognised by this region of the transposase (data not shown). The innermost copies of the 12 bp sequence can be extended by 3 bp to TAGTTAGCTGCATCG, raising the possibility that transposase may bind more strongly to this extended sequence. We have tried to determine the transposase

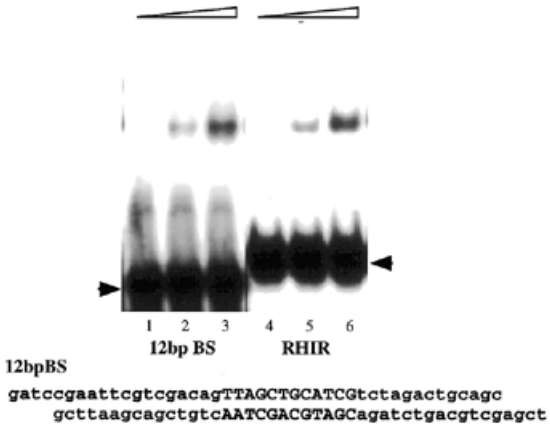

Figure 7. A 12 bp sequence is the transposase binding site. Two ^{32}P-labelled probes each containing a single copy of the 12 bp transposase binding site were incubated with 0, 1 or 3 µg of purified GST:N74 before being separated on a polyacrylamide gel. The probes were as follows: lanes 1–3, a fragment with the sequence indicated below the autoradiogram (upper case indicates the binding site); lanes 4–6, a fragment comprising nucleotides 2079–2121 of *pogo*; this contains one copy of the binding sequence TTAGCTGCCTCG as indicated in Figure 6. The positions of the free probe fragments are indicated by arrowheads.

binding site directly in footprinting experiments but for reasons that are unclear have been unable to do so.

DISCUSSION

All transposases investigated so far recognise sequences at the ends of the corresponding transposable element, and this is probably essential for transposition. These binding sites are often in the terminal inverted repeats of the element (7,23,29) although this is not always the case. The transposase of the *Ac* element of maize binds to sequences both within and adjacent to the terminal inverted repeats (32), while the binding site for the transposase of the *P* element of *Drosophila* is adjacent to the terminal repeats (33).

The DNA binding domains of all eukaryotic transposases that have been studied so far lie in the N-terminal part of the protein (22,23,32,34). The transposases of *Tc1* and *Tc3* have been shown to have bipartite DNA binding domains with the N-terminal part being responsible for recognising sequences within the terminal inverted repeats of these elements. Computer analysis of these regions indicates that HTH motifs may be responsible for sequence-specific binding and this has been confirmed by analysis of crystals of the 65 N-terminal residues of *Tc3* transposase and an oligonucleotide containing its binding site (24). This fragment of the protein forms three α-helices, the second and third of which form an HTH that binds a 20 nucleotide sequence that had previously been identified as the transposase binding site (23). This starts 12 bp from the ends of the terminal inverted repeats. Three residues in the recognition helix contact bases in the binding site. These are arginines at the positions 1 and 5 in this helix and a histidine at position 2.

We have demonstrated that *pogo* transposase, which is distantly related to the transposases of the *Tc1/mariner* family of elements, binds specifically to the ends of a *pogo* element and that its DNA binding domain is contained within the N-terminal 74 residues of the protein. This region is predicted to contain an HTH motif. We have investigated whether residues within the putative HTH are required for DNA binding by testing the effect on binding of

changing residues that might be expected to be required for recognition of the target site as well as residues that might not. Changing arginine residues in positions 1 and 6 of the second helix to alanine (R39A and R44A) severely reduced the ability of the GST:N74 fusion protein to bind to the left hand end of *pogo*, whereas changing the lysine at position 7 in the first helix (K34A) had little if any effect. In contrast, introduction of a proline residue at position 5 of the first helix (C32P) greatly reduced binding as might be expected for a change that would be expected to significantly distort an HTH. Changing the lysine immediately beyond the predicted end of the recognition helix (K48A) or the arginine 15 residues further downstream (R63A) also had no detectable effect on binding.

The protein formed by fusing GST to the N-terminal 59 residues of *pogo* transposase, GST:N59, did not bind to the left hand end of the element. This may be because the structure of the DNA binding domain that we have identified is unstable in GST:N59 rather than because there are residues essential for binding between positions 59 and 74. A similar result has been found for *Tc3* transposase. A fragment comprising the N-terminal 54 residues of this protein does not bind DNA, even though the sequence-specific binding domain of the transposase is contained entirely within this sequence (23,24).

The first of the three α-helices in the N-terminal 65 residues of *Tc3* transposase is involved in dimerisation of these fragments within the crystal (24), although it is not known if this is the case for the complete protein in solution. We have no evidence to indicate whether or not the equivalent region of *pogo* transposase is involved in binding. No stable α-helix has been predicted for this region, but neither was the first helix of the *Tc3* sequence (25).

We have identified a 12 bp binding sequence for *pogo* transposase. Two copies of this sequence are present at each end of the element, the outermost lying within the terminal inverted repeats (Fig. 6). Each copy is recognised by transposase. We expect that the binding sites nearest the ends of the element will be essential for transposition. Transposase monomers could bind at these sites by their N-termini leaving their C-terminal catalytic domains oriented towards the junction between *pogo* and flanking DNA that they must cleave to allow transposition to proceed. The inner copies of the transposase binding site may also be required for transposition. This might be similar to bacteriophage *Mu* for which there are three transposase (MuA) binding sites at each end. During transposition, a synaptic complex is formed comprising a tetramer of transposase monomers that interact with two of the binding sites at the right hand end and one at the left. *Tc3* also has two copies of its transposase binding site within each of its 462 bp terminal inverted repeats, but only the outer sites are required for transposition (van Luenen and Plasterk cited in 24). This is perhaps not surprising as the inner sites are ~200 bp from each end (35).

Pogo transposase is a member of a family of proteins with related amino acid sequences. Other members of the family include the putative transposases of transposable elements including *Pot2*, *Fot1* and *Fcc1* (12,36,37) and the products of chromosomal genes including *PDC2*, *RAG3* and *jerky* (17–19). Each of these putative transposases contains a potential HTH corresponding to that of *pogo* and these are likely to be involved in recognising the ends of the elements concerned. The proteins that are not transposases do not contain this motif, and the regions of *pogo* transposase with which they show sequence similarity lie elsewhere in the protein.

ACKNOWLEDGEMENTS

We are grateful to I. Clark and A. Dawson for careful reading of the manuscript. H.W. held a Postgraduate Studentship from The Darwin Trust and this work was supported by a project grant from the Wellcome Trust (042192/Z/94/PMG/JC/YJ3).

REFERENCES

1 Finnegan,D.J. (1989) *Trends Genet.*, **5**, 103–107.
2 Mizuuchi,K. (1997) *Genes Cells*, **2**, 1–12.
3 Grindley,N.D.F. and Leschziner,A.E. (1995) *Cell*, **83**, 1063–1066.
4 Vos,J.C., De Baere,I. and Plasterk,R.H.A. (1996) *Genes Dev.*, **10**, 755–761.
5 van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1994) *Cell*, **79**, 293–301.
6 Kaufman,P.D. and Rio,D.C. (1992) *Cell*, **69**, 27–39.
7 Lampe,D.J., Churchill,M.E.A. and Robertson,H.M. (1996) *EMBO J.*, **15**, 5470–5479.
8 Tudor,M., Lobocka,M., Goodell,M., Petit,J. and O'Hare,K. (1992) *Mol. Gen. Genet.*, **232**, 126–134.
9 Daboussi,M.-J., Langin,T. and Y.,B. (1992) *Mol. Gen. Genet.*, **232**, 12–16.
10 Levis,C., Fortini,D. and Brygoo,Y. (1997) *Mol. Gen. Genet.*, **254**, 674–680.
11 Smit,A.F.A. and Riggs,A.D. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 1443–1448.
12 Robertson,H.M. (1996) *Mol. Gen. Genet.*, **252**, 761–766.
13 Ruvolo,V., Hill,J.E. and Levitt,A. (1992) *DNA Cell Biol.*, **11**, 111–122.
14 Yuan,J., Finney,M., Tsung,N. and Horvitz,H.R. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 3334–3338.
15 Collins,J.J. and Anderson,P. (1994) *Genetics*, **137**, 771–781.
16 Earnshaw,W.C., Sullivan,K.T., Machlin,P.S., Cooke,C.A., Kaiser,D.A., Pollard,T.D., Rothfield,N.F. and Cleveland,D.W. (1987) *J. Cell Biol.*, **104**, 817–829.
17 Toth,M., Grimsby,J., Buzaki,G. and Donovan,G.P. (1995) *Nature Genet.*, **11**, 71–75.
18 Hohmann,S. (1993) *Mol. Gen. Genet.*, **241**, 657–666.
19 Prior,C., Tizzani,L., Fukuhara,H. and Wesolowski-Louvel,M. (1996) *Mol. Microbiol.*, **20**, 765–772.
20 Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 942–946.
21 Robertson,H.M. (1995) *J. Insect Physiol.*, **41**, 99–105.
22 Vos,J.C. and Plasterk,R.H. (1994) *EMBO J.*, **13**, 6125–6132.
23 Colloms,S.D., van Luenen,H.G.A.M. and Plasterk,R.H.A. (1994) *Nucleic Acids Res.*, **25**, 5548–5554.
24 van Pouderoyen,G., Ketting,R.F., Perrakis,A., Plasterk,R.H.A. and Sixma,T.K. (1997) *EMBO J.*, **16**, 6044–6054.
25 Pietrokovski,S. and Henikoff,S. (1997) *Mol. Gen. Genet.*, **254**, 689–695.
26 Clubb,R.T., Schumacher,S., Mizuuchi,K., Gronenborn,A.M. and Clore,G.M. (1997) *J. Mol. Biol.*, **273**, 19–25.
27 Schumacher,S., Clubb,T.R., Cai,M., Mizuuchi,K., Clore,G.M. and Gronenborn,A.M. (1997) *EMBO J.*, **16**, 7532–7541.
28 Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **166**, 1–19.
29 Vos,J.C., van Luenen,H.G.A.M. and Plasterk,R.H.A. (1993) *Genes Dev.*, **7**, 1244–1253.
30 Kunze,R. and Starlinger,P. (1989) *EMBO J.*, **8**, 3177–3185.
31 Brennan,R.G. and Matthews,B.W. (1989) *J. Biol. Chem.*, **264**, 1903–1906.
32 Becker,H.-A. and Kunze,R. (1997) *Mol. Gen. Genet.*, **254**, 219–230.
33 Kaufman,P.D., Doll,R.F. and Rio,D.C. (1989) *Cell*, **59**, 359–371.
34 Lee,C.C., Mul,Y.M. and Rio,D.C. (1996) *Mol. Cell. Biol.*, **16**, 5616–5622.
35 Collins,J., Forbes,E. and Anderson,P. (1989) *Genetics*, **121**, 47–55.
36 Kachroo,P., Leong,S.A. and Chattoo,B.B. (1994) *Mol. Gen. Genet.*, **245**, 339–348.
37 Panaccione,D.G., Pitkin,J.W., Walton,J.D. and Annis,S.L. (1996) *Gene*, **176**, 103–109.